CULTURE AND THE CONTAGION OF CONFLICT: SOCIAL SCIENCE AND COMPUTATIONAL APPROACHES

Michele Gelfand
MARYLAND UNIV COLLEGE PARK

08/13/2015
Final Report

DISTRIBUTION A: Distribution approved for public release.
# ABSTRACT

We witness on a daily basis conflicts which spread from individuals quickly across groups, from the highly publicized incident that occurred when the Danish daily newspaper Jyllands-Postan published an article entitled “Muhammed ansigt” which led to hundreds of protests and an escalation of violence, to the spread of conflict in Rwanda that caused the death of 800,000 individuals. This grant combined the use of social science and computational modeling techniques to illuminate the evolution of conflict contagion. We theorized that collectivism is a key driver of conflict contagion across social networks and across time through its impact on ingroup and outgroup entitativity. Our laboratory, field, and computational research showed strong support for the theory and illuminated important new scientific and practical insights. Our work was featured in top tier scientific outlets such as the Proceedings of the Royal Society B, Journal of Applied Psychology, Psychological Science, and Journal of Experimental Psychology, among others. A workshop on cultural evolution and conflict conducted through the grant has resulted in a new society for the study of cultural evolution which now has over 1000 members from many scientific disciplines.
INSTRUCTIONS FOR COMPLETING SF 298

1. REPORT DATE. Full publication date, including day, month, if available. Must cite at least the year and be Year 2000 compliant, e.g. 30-06-1998; xx-06-1998; xx-xx-1998.

2. REPORT TYPE. State the type of report, such as final, technical, interim, memorandum, master's thesis, progress, quarterly, research, special, group study, etc.

3. DATES COVERED. Indicate the time during which the work was performed and the report was written, e.g., Jun 1997 - Jun 1998; 1-10 Jun 1996; May - Nov 1998; Nov 1998.

4. TITLE. Enter title and subtitle with volume number and part number, if applicable. On classified documents, enter the title classification in parentheses.

5a. CONTRACT NUMBER. Enter all contract numbers as they appear in the report, e.g. F33615-86-C-5169.

5b. GRANT NUMBER. Enter all grant numbers as they appear in the report, e.g. AFOSR-82-1234.

5c. PROGRAM ELEMENT NUMBER. Enter all program element numbers as they appear in the report, e.g. 61101A.

5d. PROJECT NUMBER. Enter all project numbers as they appear in the report, e.g. 1F665702D1257; ILIR.

5e. TASK NUMBER. Enter all task numbers as they appear in the report, e.g. 05; RF0330201; T4112.

5f. WORK UNIT NUMBER. Enter all work unit numbers as they appear in the report, e.g. 001; AFAPL30480105.

6. AUTHOR(S). Enter name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. The form of entry is the last name, first name, middle initial, and additional qualifiers separated by commas, e.g. Smith, Richard, J, Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES). Self-explanatory.

8. PERFORMING ORGANIZATION REPORT NUMBER. Enter all unique alphanumeric report numbers assigned by the performing organization, e.g. BRL-1234; AFWL-TR-85-4017-Vol-21-PT-2.

9. SPONSOR/MONITOR'S ACRONYM(S). Enter, if available, e.g. BRL, ARDEC, NADC.

10. SPONSOR/MONITOR'S REPORT NUMBER(S). Enter report number as assigned by the sponsoring/monitoring agency, if available, e.g. BRL-TR-829; -215.

11. DISTRIBUTION/AVAILABILITY STATEMENT. Use agency-mandated availability statements to indicate the public availability or distribution limitations of the report. If additional limitations/ restrictions or special markings are indicated, follow agency authorization procedures, e.g. RD/FRD, PROPIN, ITAR, etc. Include copyright information.

12. SUPPLEMENTARY NOTES. Enter information not included elsewhere such as: prepared in cooperation with; translation of; report supersedes; old edition number, etc.

13. ABSTRACT. A brief (approximately 200 words) factual summary of the most significant information.

14. SUBJECT TERMS. Key words or phrases identifying major concepts in the report.

15. SECURITY CLASSIFICATION. Enter security classification in accordance with security classification regulations, e.g. U, C, S, etc. If this form contains classified information, stamp classification level on the top and bottom of this page.

16. LIMITATION OF ABSTRACT. This block must be completed to assign a distribution limitation to the abstract. Enter UU (Unclassified Unlimited) or SAR (Same as Report). An entry in this block is necessary if the abstract is to be limited.
In 1964, Jedu’a Abu-Sulb, a member of a Negev Bedouin tribe, became involved in a dispute during which he killed a man from the Tawara group in self defense. For several years after this, he lived in fear of revenge from the Tawara group. During this time, he married and had a son, Ayub. When Jedu’a died, the blood dispute between Jedu’a and the Tawara group transferred to his son, who now bears the burden of retaliation from a group harmed by this father.

(Ginat, 1997).

The case of Jedu’a Abu-Sulb clearly illustrates the process of conflict contagion wherein conflicts between two disputants rapidly spread across networks and time. In this case, the original dispute between Abu-Sulb and one Tawara member spread to other Tawara members via the effect that the harm had on other individuals in the group. Then, it spread further to involve any member of the Abu-Sulb’s group, including future generations such as Abu-Sulb’s son. Conflict contagion episodes like this indeed can be seen worldwide, from the highly publicized incident that occurred when the Danish daily newspaper Jyllands-Postan published an article entitled “Muhammeds ansigt” which led to hundreds of protests and an escalation of violence across the Muslim world in 2006, to the spread of conflict that transpired in Rwanda in 1994 wherein 800,000 Rwandans were killed, approximately 20% of the nation’s population (Grant, 2010). Understanding the mechanisms that produce these contagion processes is critical for both psychological theory (which tends to look at conflict in isolated episodes), as well as practice, in order to develop interventions to reduce the spread of disputes with such catastrophic consequences.

This grant is filling this void through the synergistic use of social science and computational modeling techniques which together seek to illuminate when and why conflict contagion occurs. We theorize that collectivism (and in particular, vertical collectivism, Triandis & Gelfand, 1998) is a key driver of conflict contagion across social networks and across time.
through its impact on three different types of entitativity: (1) ingroup entitativity, (2) outgroup entitativity and (3) transgenerational entitativity. In particular, when the collective self is activated, it results in higher ingroup entitativity, wherein group members are depersonalized undifferentiated entities; higher outgroup entitativity, wherein the outgroup is perceived as a unified whole whose members are perceptually undifferentiated from each other and are depersonalized entities (Kashima et al., 2005); and higher transgenerational entitativity, wherein one’s ingroup transcends past and future generations. Transgenerational entitativity can be thought of as perceptions of ingroup entitativity or interchangeability across generations (Kahn, 2010).

**Figure 1. Model of Collectivism and Conflict Contagion across Groups and Generations**

- **A. Offending Party**
- **B. Offended Party**
- **D. Outgroup Member**
- **F. Outgroup Member (1st Gen…)**
- **E. Ingroup Member (1st Gen…)**

**Amplifiers:** Public insult, Uncertainty, Threat, Degree of Violation, Significance of identity

**Propositions.** Line 1 in the above figure first illustrates the implication of collectivism and ingroup entitativity for the spread of disputes. An offense against any ingroup member is experienced as personally relevant (i.e., as if it had happened to oneself) and emotionally distressing (Lickel, Miller, Stenstrom, Denson, & Schmader, 2006; Stenstrom, Lickel, Denson, & Miller, 2008; Yzerbyt, Dumont, Wigboldus, & Gordijn, 2003). Moreover, high ingroup entitativity based on shared identity drives ingroup observers to retaliate (Lickel et al., 2006) and punish an outgroup perpetrator to regain personal and group honor. Such retaliatory behavior is not only a personal desire but also institutionalized as an appropriate response to protect the group (e.g., is endorsed collectively as a descriptive norm; Chiu, Gelfand, Yamagishi, Shteynberg, & Wan, 2010; Shteynberg, Gelfand, & Kim, 2009; Vandello, Cohen, & Ransom, 2008). Further, due to such strong group norms in collectivistic cultures, altruistic behavior toward ingroup members is particularly critical for maintaining one’s reputation as a good group member and for maintaining the safety of the ingroup and warding off future attacks from other groups (Bernhard, Fischbacher, & Fehr, 2006). Importantly, according to this perspective, an interpersonal offense develops into a system of back-and-forth intergroup revenge because people not only personally believe it is important to vicariously punish, but also perceive that others in the group expect them to do so.
Line 2 illustrates the implication of collectivism and *outgroup entitativity* for the spread of disputes. Outgroup entitativity plays a central role in collective blame and responsibility (Denson, Lickel, Curtis, Stenstrom, & Ames, 2006; Lickel et al., 2006; Lickel, Schmader, & Hamilton, 2003). Due to perceptions of outgroup entitativity, the original victim of a conflict in collectivist groups may render *any* outgroup member (even if he/she did not commit the offense) to be responsible for the offense, and consequently, to become a justifiable target of retaliation. Moreover, Line 3 illustrates the interactive effects of collectivism and both *ingroup* and *outgroup entitativity* for the spread of disputes, and in particular, how collectivism allows for the continuation of conflict even in cases in which the revenge-seeking ingroup member and the target outgroup member were not involved in the original conflict. During vicarious retribution (Lickel et al., 2006; Stenstrom et al., 2008), in which neither the person exacting revenge nor the outgroup target of revenge were directly involved in the precipitating dispute, ingroup identification and outgroup entitativity work together in concert to motivate revenge by a previously uninvolved ingroup member against a previously uninvolved outgroup member. Harm caused to one’s group becomes one’s own (ingroup entitativity) and avenging one’s own and group’s honor with retaliation against *any* outgroup member (outgroup entitativity) is personally and collectively valued and is a logical part of this cultural system. Importantly, we theorize that such processes occur even if the innocence of bystanders is known (e.g., they were not involved, nor could they have prevented the original act (sins of omission or commission; Lickel et al., 2003)). Put differently, contagion to restore individual and group honor is blind to guilt or innocence of outgroup bystanders in this process.

Line 4 illustrates the dynamics of contagion of conflicts across generations in collectivistic cultures. Due to greater *transgenerational entitativity* (i.e., the belief that one’s ingroup transcends past and future generations, TGE) collectivism makes it more likely that future generations of ingroup members, who did not witness the original act, will have biased memories of conflicts that occurred in previous generations, and will feel obligated to retaliate on behalf of previous ingroup generations. In addition, because one’s ingroup transcends future generations, TGE may relate to self-sacrificial behaviors for the benefit of restoring the group’s honor for previous and future group members. We note that such behavior is not only fueled by a personal desire but is also institutionalized as an appropriate response to protect the group (e.g., is endorsed collectively as a descriptive norm).

Finally, it is worth noting that the very processes that account for conflict contagion may also promote the *spread of forgiveness*. In collectivistic cultures, responsibility to apologize reaches a far greater web of actors and includes the collective as a whole (Maddux & Yuki, 2006). Representative group members (e.g. senior leadership) who have no personal guilt, or even involvement, often apologize on behalf of the group (Greenberg & Elliot, 2009), and these indirect apologies are especially common in collectivistic cultures (Chiu & Hong, 1992; Zemba, Young, & Morris, 2006). There may be a greater expectation, and willingness, to apologize on behalf of ingroup members (i.e., ingroup entitativity, Line 2) to outgroup victims and outgroup bystanders (i.e., outgroup entitativity, Line 3) in collectivistic groups when one’s ingroup member has offended the outgroup. Furthermore, there may be a greater willingness to accept apologies that are given by outgroup perpetrators and bystanders who are contemporaneous and distal to the conflict in collectivistic cultures.
Our theory also suggests that there are likely numerous situational factors that moderate the extent to which conflict escalates. Put simply, conflict contagion is dynamic and subject to situational effects. Line 5 illustrates several factors that might amplify cultural differences in conflict contagion. First, situations that cause people to engage in automatic processing and rely on well-learned cultural tendencies are theorized to exacerbate conflict contagion in collectivistic groups. For example, situations which increase the salience of cultural values and norms may cause conflicts to be more contagious in collectivistic groups. To the extent that cultural values and group norms are reinforced through peer expectations (Chiu et al., 2010; Shteynberg et al., 2009), they are made more salient when conflicts are in public wherein harm to one’s ingroup is being observed by others, as compared to when they happen in private. Accordingly, we would expect that conflict contagion processes are exacerbated in contexts where offenses are public and less so when they are private. Situations in which there is high threat and uncertainty activate strong epistemic needs for individuals to identify with groups as epistemic authorities and conform to group norms (Webster & Kruglanski, 1994). Thus, we would expect that such factors will amplify cultural differences in the above processes. That is, when people face a high degree of threat they strongly hold on to their cultural identities in order to reduce anxiety (Greenberg, Solomon, & Pyszczynski, 1997). Therefore, we expect that individuals facing uncertainty and group threat—be it situational (Hogg, Meehan, & Farquharson, 2010) or an individual difference (e.g., need for closure, Webster & Kruglanski, 1994; self-concept uncertainty, Mullin & Hogg, 1998)—should show stronger reliance on entitativity and play a more pronounced role in the transmission of conflict across networks and time.

**Experimental and Qualitative Interviews of Conflict Contagion**

We have theorized that conflict has the potential to be more contagious in collectivistic cultures. Our research, published in the *Proceedings of the Royal Society B* (Gelfand et al., 2012) illustrated support for the theory with qualitative interviews across eight nations. In this study, structured interviews were conducted in Egypt, Iraq, Jordan, Lebanon, Pakistan, Turkey, UAE, and US. The Pakistani and Middle Eastern samples were of particular interest because they constitute a type of collectivism in which group members are expected to sacrifice self interests for the group, and there is a sharp demarcation between the ingroup and outgroups. A total of 184 participants—composed of community members varying in age, gender, socioeconomic status, and rural-urban residency—were interviewed across all countries. The researchers asked interviewees to talk about the interrelationship between their honor and honor loss and others’ honor and honor loss. These questions included: 1) *Is your honor (sharaf) related to the honor (sharaf) of other people, and whom? How does something affecting your Sharaf affect the Sharaf of others?* 2) *Likewise, does the loss of honor of others affect your honor?* 3) *Whose honor is most important to you?* 4) *How does it affect you?*

We conducted both qualitative and quantitative analyses of responses to these questions. Using analyses of word frequency (LIWC; Pennebaker, Francis, & Booth, 2001), we examined the extent to which people discussed a wide range of social entities that are involved in the contagion of honor loss. An overall Social Index was calculated for each interviewee as a percentage of the total word count of the interviewee’s responses to all questions. This Social Index included family members, with both social entities in the nuclear family (e.g., spouse, parents, children, siblings) and social entities in the extended family (e.g., aunts, uncles, cousins,
relatives, ancestors); non-family relationships such as friends, coworkers, classmates, neighbors, and groups that comprise an extended network of social ties (e.g., neighborhood, village, tribe, company, and university); and large-scale social identity groups, such as one’s nationality, ethnicity, religion, and abstracted groups, including civilization, society, and culture.

Findings from this study illustrated a clear and re-occurring theme of the interchangeability of honor and contagious effect of honor harm across the Middle East (ME) and Pakistan as compared to the U.S. Middle Eastern participants as a group mentioned more social entities than did Americans, showing that the “web” of people to whom one’s honor is related is much wider in these countries compared to the US. On average, Americans mentioned social entities in 3.34% of their responses, while the ME and Pakistan countries mentioned 7.53%, with interviewees from Jordan and Iraq scoring as high as 11.67% and 10.14%, respectively.

Qualitative examination provided a richer account of cultural differences in the degree to which one’s honor gain and loss is interrelated to the gain and loss of others’ honor. Responses from US respondents tended to differentiate one person’s honor from another’s. Overall, Americans respondents did not think that their honor loss would affect the honor of those around them. One respondent stated: “People might look at my wife, a little less friendly. But yet, they shouldn’t really, I mean, if it’s my issue, not hers”. Another American interviewee explained “The fact that I know them? Um it shouldn’t. I would hope it wouldn’t…I believe honor is each person, you gotta look at each person individually”. In rare cases where a person’s honor was related to another’s, American respondents included a small circle of people to whom their honor is related: “My values and honor was probably established by my upbringing with my parents. My mom um, but it’s not related to anybody else”. Furthermore, American respondents discussed being less impacted personally by others’ honor loss, noting in particular that it would not impact their own honor: “it would affect me…but it wouldn’t affect my honor, no”. Another interviewee stated, “[I would] probably feel bad for them, I would be upset, but I wouldn’t lose my mind over that”. Others noted that they would want to help others in honor loss situations (e.g., “If they go through a hard time where they don’t have honor at school anymore, I’m going to try and fix it”); yet, others’ honor loss would be much less contagious to one’s own sense of honor among American interviewees.

The high entitativity among collectivistic group members would suggest that the honor of an ingroup member is interchangeable with that of another member. As predicted, ME and Pakistani respondents frequently discussed the interchangeability of honor. One UAE interviewee explained, “[Yes], members of my family, my extended family, my people…their honor is related to mine because they are members of my family. What touches me touches them and what touches them touches me”. An interviewee from Egypt similarly commented that “Of course my honor is my husband’s honor, my children’s honor. All of us are one, the honor of any one of us is the honor of the other”. Lebanon interviewee echoed this sentiment by explaining, “The word honor in and of itself carries a non-individualist meaning…it’s effects are interchangeable among family members in what is related to honor”. The contagion of honor loss can extend to larger social identity groups, including one’s religion, gender, and other generations of one’s family. For example, a Jordanian interviewee commented on the different spheres of honor loss: “Firstly his personal honor, then his children's honor and his country's
honor”. A Turkish interviewee likewise stated that his honor extended beyond the closest circle to “the society in which I belong”. One UAE interviewee summed it up, “We all live in one boat and one society; therefore a drowning person will affect the whole of social ties”.

The interchangeability between related others’ honor suggests that when a person is harmed, other individuals in the group would be similarly harmed. Indeed, responses from the ME region and Pakistan frequently alluded to the ripple effect of honor loss to other group members. Commenting on the contagion of insults, an Egyptian interviewee explained, “I am a Qadwa, from my parents, their name would be shaken, my husband’s name as well if something causes my honor to be insulted”. Beyond the immediate family, ripple effects from honor loss extend widely (“close relatives, brothers and cousins, and tribe those who relate to his honor then people who live nearby, for example the district where he resides, neighbors, his honor, and his reputation” (Iraq)) and quickly (“…if [the honor attack] is not confronted it spreads like an infection and I become ashamed” (Lebanon)). And finally, honor loss is permanent: “Honor is never forgotten and if it is harmed it can never be erased” (Jordan).

Overall, the interview data from Gelfand et al. (2012) revealed that for collectivists, honor is interchangeable, especially among one’s family and extended networks; and it is contagious—when an ingroup member is harmed, people are much more affected by it and such effects spread through a much wider network of people. These findings suggest that group members are more entitative in collectivistic groups as compared to individualistic groups, and entitativity, in turn, affects how people react to instances of a group member being harmed.

Extending this qualitative analysis, we have found evidence for cultural variation in conflict contagion in the laboratory (Gelfand and colleagues, in prep). In a free recall study, collectivists were more likely to report wanting to take revenge on behalf of a group member who was made to feel humiliated, whereas individualists actually distanced themselves from other group members when they are humiliated. We also showed behavioral evidence for the phenomenon of conflict contagion in the laboratory. Using a modified dictator game, we had individuals who varied on collectivism take part in a between-subject experiment in which they witnessed an out-group member commit a harmful act against either (1) an in-group member with whom they shared a social identity or (2) a neutral party with whom they did not share a social identity. Consistent with our predictions, participants who are higher on collectivism were more likely to punish a third party when they share a social identity with the victim but were less likely to do so when they did not share a social identity with the victim. Using a cross-cultural sample, a third study illustrated vicarious revenge as a mechanism for conflict contagion. Participants in Turkey and the US read two scenarios in which one of two friends was insulted in public by a stranger. Modeled after vignettes developed by Cohen (Cohen & Nisbett, 1997; Vandello, Cohen, & Ransom, 2008), these offenses were honor-threatening transgressions: the insults were obvious, intentional, in public, and were not followed by any apology by the perpetrator. The victim’s friend confronted the offender in a subsequent encounter. For each scenario, participants rated their reactions to the revenge according to how much they felt it was (un)necessary, (un)justified, (not) understandable, (dis)honorable, and (un)acceptable, etc, and evaluated the revenge-seeker. The expected cultural differences emerged, with revenge more sanctioned among Turkish respondents across the two scenarios. Furthermore, Turkish
respondents were more likely to believe that the revenge-seeker is similar to them. These studies will be submitted to the Journal of Experimental Social Psychology for review.

Our other research (Shytenberg et al., under revision, European Journal of Social Psychology) presents another study on collectivism and contagion that captures situational dynamics of the phenomenon. In particular, we examined how collectivism, in conjunction with other situational factors (e.g., having a higher versus lower need for closure) affects the contagion of others’ injustices. We hypothesized that people with more collectivistic attitudes are more likely to consider the treatment of a teammate or a coworker as relevant to their cognitive and behavioral reactions, particularly when they have low need for closure (i.e., engage in greater information processing and perspective taking) (Kruglanski, 2004; Kruglanski, 2009). We conducted a field and a laboratory study to test our hypothesis. In the field study, we tested our hypothesis in organizational settings with employees (and their supervisors) from a variety of companies. We also conducted a laboratory study that allowed us to manipulate the unjust treatment of a fellow teammate at the hands of an authority and then subsequently measured personal evaluations of the authority’s fairness. Both studies provided support for our hypotheses. We found that collectivism and epistemic motivations work in concert to make another’s justice one’s own. That is, the justice treatment of others has a larger influence on people who are simultaneously higher (vs. lower) in collectivism and lower (vs. higher) in the need for closure. Notably, we found that teammates’ mistreatment was not only relevant to laboratory participants’ justice judgments, but also to the turnover intentions and supervisor-directed helping behaviors of employed adults. We believe this work improves our ability to predict when the injustice of another will spread beyond the victim.

To explore other situational moderators of conflict contagion, we used an adapted/modified dictator game design (Bernhard, Fischbacher, & Fehr, 2006, Nature) in which individuals from Jordan and the U.S. came to the lab with three friends. They were randomly assigned to be in trios in which a dictator (Person A) is given a monetary endowment which he or she may choose to give away some or not at all to a recipient (Person B) who starts without an endowment. The allocation amount is revealed to a punisher (Person C) who may choose to pay some of their own starting endowment in exchange for punishing Person A at different levels. One of the added advantages of the current design is that it can be adapted to investigate various research questions by creating experimental trios that comprise various combinations of group membership. For example, for some trios, Persons B and C were friends who came to the study together. This condition can investigate how people react when the ingroup is harmed, a central research question. In other trios, Persons A, B and C are friends who came to the study together and in other trios A, B, and C are all strangers. The flexibility of the design will also offer insight into possible vicarious revenge scenarios, parallel to some of our other studies. In this case, one condition assigned participants to be in a quartet rather than a trio, in which Person C may punish Person A’s group member (Person D) rather than Person A directly. This study is currently being completed and being analyzed.

The above studies were based on one-shot interactions. Our research on this grant advanced the study of conflict to examine how conflicts get transmitted over time across groups. Based on our theory of culture and contagion research, we anticipate that conflicts are more contagious across time in collectivistic groups, with the result that conflicts persist much more
across generations, including among individuals who were not involved in original conflicts. To
test the idea that collectivistic individuals are more likely to transmit conflict information
through their narratives across time, we had previously designed and piloted a study with
guidelines set forth in previous studies using the Bartlett method of serial reproduction (Lyons &
Kashima, 2001; Lyons & Kashima, 2003). This method is applicable to study contagion
processes and the distortion of collective memory, as it has been used to understand information
transmission and collective memory for rumors and stereotypes (Lyons & Kashima, 2001; Lyons
& Kashima, 2003). Groups of four participants each completed a chain of reproductions: the first
person read a story we provided that describes a group conflict, while the other three read and
reproduced the version that was passed down to them from the previous person, akin to the way
that collective memories are spread from generation to generation (i.e., using a telephone game
metaphor). One major strength of this design is that despite the same starting point (researcher-
created story), it potentially produces a story in the end that has been transformed to include
information rich in the group-level biases of the storytellers, and lends itself to be analyzed with
different approaches, as detailed below.

We successfully completed and published a study in the Journal of Experimental Social
Psychology on the effects of ingroup conflict involvement on the types and strength of group
biases that emerge in people’s collective memories. Chains of participants received the initial
story about two groups in conflict both of whom were strangers to them (Control condition) or
one of whom were friends with the participants (Ingroup condition). We analyzed the stories
produced at the beginning of the chains by the first person and at the end of those chains by the
fourth person as a function of condition. Our investigation employed multiple statistical analyses
from various angles, and together, they revealed a remarkable level of ingroup biases in the
stories people retold as well as in their evaluations—that is, their take-home attitudes—about the
conflict and the conflict participants.

First, we analyzed the content of the stories using our internally developed Honor
Dictionary. In particular, we were interested in the use of words related to morality and
wrongdoing. Words belonging to the morality category reference qualities that contribute to the
(non)integrity of a person or group. These words include ethic, (un)fair, right, justice, virtue, etc.
Words in the wrongdoing category refer to acts of misbehavior and wrongdoing, and include
wrong, lie, rude, guilt, etc. As expected, linguistic analyses conducted on these two categories
showed increasing use of morality words within Ingroup chains compared to a consistent level in
the Control condition chains. Furthermore, use of wrongdoing words remained consistent in the
Ingroup condition as compared to the control condition in which it decreased from the first to
fourth person as information is increasingly lost in transmission down the chain. Second, we also
employed content coding analyses. Using the codebook developed last year, we coded each unit
of thought in the reproduced stories. A team of trained coders tallied the frequencies of when a
particular detail is distorted to exaggerate the blame of one of the groups versus when it is
distorted to downplay the blameworthiness of one of the groups. Examples of blame
exaggerations include stating ambiguous information as fact by omitting one group’s expressions
of denial; creating new information that implies the consistency and repetition of an isolated
event; and exaggerating either actions or consequences. Blame attenuation includes making
excuses (e.g. emphasizing extenuating circumstances) or justifications for a group’s behavior
(e.g. it was necessary; the other group deserved it); particularizing the occurrence of one’s
blameworthy behaviors; and providing information that a group showed conciliatory behaviors (e.g. making amends or reparations). Results showed that ingroup bias is greatest when it comes to *attenuating the ingroup’s blameworthiness*, suggesting that people’s reproductions are distorted to downplay the fault or blame of one’s own group.

Additional analyses examined participants’ ratings of empathy which were assessed through questions about how understandable, (in)appropriate, justified, etc., the groups’ actions were relative to each other. We found that people in the Ingroup condition showed significantly higher and increasing levels of empathy toward one’s own group, as compared to Control condition participants who showed a neutral and consistent level of empathy toward either group. In tandem with the distortions in people’s reproductions, people also showed *bias in their evaluations toward each group* if their group was involved in the conflict. Participants rated how well they thought various adjectives described each conflict group. These words encompassed both positive traits (e.g. respectable, honorable, moral) and negative traits (e.g. malicious, manipulative, cruel). Due to a drop in information between communicators down the chain, Control participants’ positive and negative trait attributions for both groups diminish across the chain; however, positive attributions toward one’s ingroup and negative attributions toward the outgroup remain stable within the Ingroup condition chains. These findings suggest that people remain steadfastly loyal to their ingroup members in spite of the circumstance that those ingroup members were equally deserving of blame originally. However, distortions in the retelling of the conflict lessened ingroup blameworthiness over time and disproportionately carried through the outgroup blameworthiness. This research is the first to examine conflict distortion across chains, with important theoretical and practical implications. It appeared in the *Journal of Experimental Social Psychology* in 2014, was quoted in the press, and was featured in the Harvard Program on Negotiation newsletter

*Computational Evolutionary Game Theoretic Modeling*

In our computational and evolutionary game theoretic, we took a two-pronged approach. On one hand, we are interested in the evolutionary basis of third-party punishment: under what conditions is third-party party punishment evolutionary stable or advantageous and how does it relate to the evolutionary dynamics of cooperation? On the other hand, we are interested in how third-party punishment and related cultural properties that distinguish collectivist and individualist groups from each other may affect the dynamics of conflict itself, both between and within groups. Necessary and important questions we needed to answer in both these approaches is how third-party punishment differs from regular punishment, how it can be implemented in an agent-based evolutionary game-theoretic model and what the implementation choices are, and how these differences and choices affect the evolutionary dynamics. We have thus developed evolutionary game models with the purpose to answer some of these basic questions that will need to be considered in more realistic models. As is the standard for regular punishment in evolutionary game models, we have considered third-party punishment as an extra action available to agents that can be taken after agents interact in a regular game-theoretic stage game that reflects a situation with the potential for conflict. Whether or not agents engage in third-party punishment is governed by a behavioral trait that may be either present or absent in an agent’s game strategy. Our work, published in the *Proceedings of the Royal Society B*, is the first to show what conditions cause the evolution of third party punishment. We have also developed
models of group-based conflict in which we explore the effects of collectivistic group-entative individuals, individualistic individuals, their effect on group conflict, and the conditions under which either collectivistic or individualistic individuals emerge. This work is now under review at Nature Scientific Reports. Each is described in turn.

Evolution of Third-Party Punishment

Third-party punishment (3PP) is different from regular punishment in several crucial ways. The most obvious difference is that agents who engage in 3PP punish agents that they themselves have not encountered in an interaction that has the potential for conflict. This simple difference already gives rise to several implementation constraints and choices. If our evolutionary game model is to allow for the potential of 3PP, then it must allow agents to encounter and be able to punish other agents that they themselves have not played in the stage game, and they must be able to acquire information about the other agent’s past interactions with other agents in order to make their choice of whether to punish or not. A simple way to model this is to have agents learn about other agent’s passed interactions with a probability i upon encountering them, similar to the implementation of notions of reputation in other evolutionary game theoretic models.

Considering these model requirements and implementation choices inherent with the investigation of 3PP, we have generated an evolutionary game-theoretic model that strives to be as simple as possible while allowing for the exploration of the effects of various important third-party-punishment related model parameters to study the evolutionary dynamics. We summarize this model in the following paragraph and then present some results on the dynamics of 3PP and the evolution of within-group cooperation. This model currently only considers within-group interactions, it is important to understand the dynamics of 3PP within groups before moving on to between-group interactions, but the model is designed to allow for the straight-forward extension to include between-group interactions once the basic dynamics with within-group interactions are understood.

Our model consists of a population of agents in which each agent belongs to a certain distinct, observable, mutually exclusive group (n = number of groups in the population). Each generation, each agent interacts with other agents in the group in a two-player stage game that presents an interaction with the potential for conflict. The choice for this stage game can be any 2x2 payoff matrix, like the Prisoner’s Dilemma, the Hawk-Dove game, etc., and the two general actions available to agents are Cooperate or Defect. Each agent’s strategy in this game is given by a behavioral trait or “gene” that can be continuous (giving a probability of cooperating) or binary (the agent either always cooperates or always defects). After the stage game interactions, agents are paired k times with other agents in the group and are given the chance to engage in 3PP. Whether or not agents will engage in third party punishment is dependent on whether the agents have the 3PP trait (modeled like the stage-game strategy trait, but separate), and whether or not they receive information about the other agent’s past behavior in the stage game, which is given with probability i – the information level. Both the pairings for stage game interactions and punishment are random for now. Punishment, as is standard, harms the punished agent by an amount ρ at a cost to the punisher l, where ρ > l. After these interactions, agents within groups reproduce according to their acquired payoffs from the stage game minus the punishment-related
harm. This is implemented through a pair-wise imitation process that is analogy to social learning, each agent considers a random other agent in the group and switches to this agents’ behavioral traits with a probability that is proportional to the payoff difference between the agents. With probability \( \mu \) agents mutate their behavioral traits. After these within-group population dynamics, each agent, with chance \( \nu \) – the migration rate, chooses a group out of the population to switch membership to, based on the average payoff of agents in the groups. The group is chosen with probability proportional to the average payoff. This payoff-based migration introduces a form of group selection to the environment, where groups that fare well attract agents while groups that do not have a tendency to loose agents. Given this simple model, we already have several model parameters (\( n, k, i, \rho, l, \nu, \mu \), the stage game payoffs) that may have a significant effect on the evolutionary dynamics of cooperation and 3PP. These effects need to be understood in order to understand the evolutionary basis of 3PP and its relationship to various environmental factors.

With the above model we have run a number of pilot experiments in order to attain a general insight into the dynamics of 3PP and cooperation. We have found that 3PP plays a crucial role in the dynamics of cooperation and defection within groups. Figure 1 shows a snapshot example of the dynamics of both 3PP and cooperation in a population. The stage game chosen was a version of the Prisoner’s Dilemma, where \( b \) is the benefit an agent receives if the other cooperates, and \( c \) is the cost of cooperating. It is clearly observable that the 3PP is related to the evolution of cooperation within the population: When 3PP rises in the population, the number of Cooperators rises soon after, but when 3PP decreases once a relatively high amount of cooperation has been established, Cooperators decrease and cooperation breaks down until third-party punishers again increase. This illustrates that 3PP and cooperation dynamics are closely related, and we are in the process of fully describing and investigating this relationship.

![Figure 1](image)

**Figure 1:** Population proportions of Cooperators and Third-Party Punishers over 1000 generations. Population initialized randomly to approximately 50% Cooperators and 50% Third-Party Punishers (independently distributed). Population consists of \( m=20 \) groups of 20 agents. Other model parameters are \( b=4, c=1, k=8, i=1, \rho=3/2, l=1/2, \nu=0.1, \mu=0.01 \).
Since the proportion of third-party punishers in the population is related to cooperation in a manner such that higher 3PP leads to higher rates of cooperation to emerge, and decreases in 3PP leads to decreases in cooperation, our model results thus far suggest that 3PP is indeed adaptive in some circumstances in the sense that it allows populations to achieve higher payoffs through higher rates of cooperation.

In order to explore the specific differences between regular punishment and 3PP and their effects on cooperation, we have compared the model with 3PP to a model with just regular punishment. Our results have shown that under a variety of parameter conditions, 3PP leads to approximately twice the average amount of cooperation in the population over time. This shows that the specific advantages of 3PP, including that many agents can punish a single defector, gives 3PP a significant advantage in promoting cooperation. These results hold for even when the punishment cost and fines of regular punishment are increased so that the total amount of fines that may be applied against punishers is equal to that in the 3PP condition, where the number of encounters \(k\) determines the maximal amount of punishment fines that can be applied.

One observation from our pilot model results is that the problem of second-order free-riding, that has been shown to exist for regular punishment, is also persistent for 3PP. This is evident because once cooperation has been established to a fairly high degree, punishers decrease, because non-punishers do not pay the cost of punishing. This leads ultimately to a short term collapse of cooperation until punishers again increase in the population in conjunction with cooperation. Hence the second-order free-rider problem induces the cyclical nature of the population dynamics in Figure 1, and the general question concerning the conditions under which 3PP can evolve to a more stable extent in populations is still unanswered.

To solve the puzzle of the evolution of 3PP, we draw on insights from the evolutionary game literature on direct punishment. Recent research showed for the first time how responsible direct punishment (punishment of non-cooperators only) can evolve even when allowing for the possibility of anti-social punishment (punishment of cooperators) (Hilbe & Traulsen 2012). The key to the evolution of responsible direct punishment in this work was the existence of punishment reputation. Accordingly, we asked: can punishment reputation also account for the evolution of responsible 3PP? Our results, discussed in more detail below, show that punishment reputation on its own cannot, leaving a puzzle concerning the conditions that lead to the evolution of 3PP.

To address this puzzle, we draw on classic sociological and psychological theory and theorize that social-structural constraints in human populations play a crucial role in enabling the evolution of responsible 3PP. The constraints we are interested in specifically are strength-of-ties (Granovetter 1983) and mobility (Oishi et al. 2007), which have been shown to have wide-ranging consequences for humans (see Oishi et al., 2010 for reviews). For centuries, humans lived in social-structural contexts characterized by strong social ties (where people interact with great frequency) and low mobility (where people are unable to exit or switch social groups with ease). Such strong social-structural constraints can be a consequence of kinship, which plays an important role in the evolution of cooperation and related behaviors (West et al. 2011, Nowak et al. 2006). It is precisely these conditions under which we anticipate 3PP to be adaptive because
punishers can more effectively induce self-interested agents to cooperate under such constraints. More specifically, only in contexts of high social-structural constraint can punishment reputation foster a culture that incentivizes self-interested agents to cooperate and hence make responsible punishment both beneficial to the individual and the collective. By contrast, in socio-structural contexts characterized by low strength-of-ties (i.e., where agents do not interact with great frequency) and/or high mobility (i.e., where they can exit the group with ease), any given individual's predisposition to punish misbehaviors will not have the same motivational force to sway self-interested agents towards cooperation, rendering responsible punishment ultimately costly to individuals and hindering the evolution of such 3PP.

To evaluate our hypothesis that these constraints are critical for the evolution of responsible 3PP, we implement variable notions of strength-of-ties and mobility in a structured population model. There exists a large evolutionary game literature exploring effects of population structure on evolutionary outcomes (see Nowak et al. 2010, Shakarian and Roos, 2012), but structured population models on punishment have only considered direct punishment and not 3PP. Our model results show that, when other mechanisms alone are unable to, responsible 3PP can evolve and induce cooperation in structured populations with the help of punishment reputation. However, high strength-of-ties and low mobility are critical for this process. When responsible 3PP evolves, it does so as an ultimately non-altruistic trait. The behavior acts as a signal to potential co-players in the neighborhood that non-cooperation will not be tolerated. High strength-of-ties and low mobility allow clustered agents engaging in responsible 3PP to induce cooperation in their neighborhood. By inducing such local cooperation, clusters of 3PP agents increase their own payoff and spread. This process leads to the emergence of responsible 3PP in the population as a whole. By contrast, low strength-of-ties and high mobility prevent clusters of 3PP agents from inducing a local culture of cooperation, and hence responsible 3PP does not evolve. This work is also of the first to illuminate the conditions under which 3PP evolves while allowing for non-responsible punishing strategies.

In order to study the effects of strength-of-ties and mobility on the evolution of 3PP, we extend our above model to include population structure by placing agents on a graph, following the large literature on spatially structured evolutionary games, and pairings for the game interaction and punishment opportunities can only occur between agents that are connected on the graph. A complete graph, where all agents are connected to all others, is the equivalent of a well-mixed or non-structured population.

Strength-of-Ties: In his classic Sociological work, Granovetter (Granovetter 1983s) measured tie strength between two humans in terms of how often they interacted with each other during a period of time. Since our model assumes that in general each agent has an equal number of interactions in a given time period, this means that in a given time period or generation, an agent with few connections has a relatively high number of interactions with its few neighbors, while an agent with many connections has a relatively low number of interactions with a greater variety of agents. Thus, by Granovetter's definition, the former agent has high strength-of-ties whereas the latter has low strength-of-ties. The degree of a node is hence directly inversely correlated with the associated agent’s average strength-of-ties. (Note that if agents were paired with all their neighbors for interaction in each generation -- as is often done in the evolutionary game literature -- the concept of strength-of-ties would be eliminated, as all agent pairs would
have an equal number of interactions in any given time period.) We shall denote the average strength-of-ties in a population as $1/d$, where $d$ is the average node degree of the graph representing the population structure. Since a complete graph has the highest possible average degree, a non-structured or well-mixed population of size $n$ has the lowest strength-of-ties possible $1/n$.

**Mobility:** As a conceptual replication we also explore a second form of social-structural constraint: residential mobility (Oishi et al. 2007). Residential mobility is the degree to which humans are able to change their location, and, as a result, their position within the social network within a population. Some human populations, particularly those that are individualistic, have very high mobility where people can easily exit the group, whereas others, particularly collectivistic cultures, are much more dependent on others and are less able to easily exit the group (Oishi et al. 2010, Schug et al. 2010). In mobile populations, humans may change their location for a multitude of reasons. We implement a simple model of the concept of residential mobility using a probability $m$ with which, at the beginning of each generation, an agent switches position with a randomly chosen other agent in the population.

A complete strategy determining an agent's actions in this environment consists of a strategy for the cooperation game phase and one for the punishment phase. In the cooperation game phase, there are three possible strategies: (C)ooperate, (D)eject, and one an (O)pportunistic strategy. Cooperators and Defectors simply always cooperate or defect, respectively. Opportunistic agents take the punishment reputation of neighbors in account when deciding to cooperate or defect in the cooperation game phase. We assume punishment reputation always exists, i.e. agents know the punishing strategies of their neighbors. Opportunistic agents choose the action that gives them the higher expected payoff given this information. In the punishment phase, there are four possible strategies that may condition the decision to punish or not on the action of the other agent in their cooperation game: agents can punish (R)esponsibly (only punish Defectors), (A)ntisocially (only punish Cooperators), (S)pitefully (punish indiscriminately), or they can be (N)on-punishing (punish no-one).

Our model results show that the evolution of responsible 3PP critically depends on conditions of high social-structural constraint, i.e. high average strength-of-ties and low mobility. Figure 2 plots the average long-term proportion of responsible punishers in the population under a) varying strength-of-ties and b) varying mobility. We vary strength-of-ties by structuring the population on graphs of different average node degree. In order to use population structures with realistic social-network characteristics, we used Watts-Strogatz small-world networks (Watts 1998): each agent is connected to $d$ nearest neighbors on a ring and then each edge (holding one end fixed) is reattached to a random node with probability 0.1, giving average strength-of-ties $1/d$. The degree of mobility is varied through our mobility parameter $m$. Populations were initialized with all Opportunistic Defectors and Non-Punishers. We can observe that conditions of high social-structural constraint, i.e. high average strength-of-ties and low mobility enable the evolution of responsible 3PP. The higher the strength-of-ties, and the lower mobility ($m$), the easier it is for responsible punishment to evolve and be sustained at high population proportion in the population. The benefit of cooperation $b$ quantifies the effectiveness of cooperation: the lower $b$, the more difficult it is for cooperation and responsible punishment to evolve. The rate of cooperation (percentage of cooperative actions) throughout these simulations is virtually
identical to the proportion of third-party punishers in the population, hence only this quantity is shown.

Figure 2: Surface plot of long-term average population proportions of responsible 3PP under varying constraint conditions. The z-axis (height) shows the long-term average proportion of responsible 3PP in populations under varying b and average tie-strength (part a) or mobility rate (part b). Higher Locations (lighter colors) mean higher population proportion. Cooperation rates (not shown) are virtually equivalent to the proportion of responsible 3PP. Total population sizes are 996 agents in order to allow for an equal number of all types. Results are similar when populations are initialized with all Opportunistic Defectors and Non-Punishers. Long run average proportions were attained from averaging 100 simulation runs over 5000 generations for populations of with model parameters c=l=1, ρ=3, µ =0.01. For part a) m=0, for part b) d=4.

Figure 3 shows representative evolutionary trajectories for single simulation runs under high strength-of-ties sufficient for the evolution of 3PP (Figure 3a) and under low strength-of-ties not sufficient (Figure 3b). Under high strength-of-ties, responsible 3PP quickly invades the population and remains the prominent punishment strategy, while under low strength-of-ties, non-punishers and even anti-social punishers comprise the prominent punishment strategies. Again, the percentage of cooperative actions in a population closely approximates the percentage of responsible punishers in the population at that time.
Figure 3: Typical evolutionary trajectories for single model simulation runs under high strength-of-ties that enable and low strength-of-ties that prevent the evolution of responsible 3PP. For readability, the plots show the aggregated proportion of punishment phase (top panel) and cooperation game phase (bottom panel) strategies over time separately. The lower panel also shows the average cooperation rate (percentage of cooperative actions) in black. Model parameters are $b=4$, $c=l=1$, $\rho=3$, $\mu=0.01$. Populations are 1000 agents and initialized with all Opportunistic Defectors and Non-Punishers. Populations are structured on Watts-Strogatz networks of $d=4$ (a) and $d=14$ (b), giving average strength-of-ties $1/4$ and $1/14$, respectively.

As an example of how responsible 3PP can induce cooperation and proliferate, see the illustration in Figure 4, a small part of a network under different configurations showing how responsible (R) 3PP affects local payoffs. A lone R punisher, as shown in the top-most configuration, is not enough to induce cooperation and actually suffers relative payoff loss compared to neighbors. However, if the R punisher is joined by another R punisher (e.g., see middle configuration) in the neighborhood, together they can induce cooperation, gain a large relative payoff advantage, and hence be likely to spread. If the payoff advantage allows the R punishers in the neighborhood to increase in number, the relative payoff advantage can become even greater (e.g. see bottom-most panel). Put simply, a R punisher increases the likelihood that nearby R punishers will be able to induce cooperation in their co-players. Hence, the agent
promotes the existence of other local R punishers that in turn encourage local cooperation further. Through this, responsible 3PP and cooperation can spread throughout the population as a whole.

Figure 4: Examples of a small part of a network under different configurations of punishers, showing how the existence of responsible 3PP agents affects local payoffs and cooperation.
All nodes are assumed to be opportunistic and the node labels designate the punishment phase strategy: R=Responsible, N=Non-Punishing. Non-labeled nodes are assumed to be non-punishing. Blue nodes choose to cooperate based on the punishment reputation of neighbors, red nodes defect. Expected payoff calculations are shown next to the nodes. E.g. in the middle configuration, the top-most right agent defects because $c/\rho > 1/4$ (see Eq. 1). The agent has an expected payoff of $b/2 - \rho/4 - \lambda/2 = 0.75$ because it has a $2/4$ chance of being paired with a cooperating agent in the game phase, giving $b$, and then a $1/4$ chance of being paired with an R agent who will punish them by $\rho$, and a $2/4$ chance of being paired with an agent who defected and thus it will punish at a cost $l$.

Relevant model parameters are $b=4$, $c=l=1$, $\rho = 3$.

Responsible 3PP does not evolve in well-mixed populations, which have the lowest possible strength-of-ties. Thus the availability of punishment reputation and the ability of opportunistic agents to take this information into account in their cooperation game decision are not sufficient for the evolution of responsible 3PP. Punishing responsibly as a third party is only ultimately beneficial to agents, and hence can spread, if there exist enough other responsible punishers to induce cooperation in potential co-players. If there are not enough other responsible punishers, punishing responsibly is a wasteful and ultimately costly act to the punisher; non-punishers would have a payoff advantage and quickly begin invading the population (the second order free-rider problem). In an unstructured or well-mixed population, as we shall elaborate below, arriving at a state where there exist sufficient responsible punishers is extremely difficult to achieve from a population of non-Punishers.
The fact that high social-structural constraint enables the evolution of responsible 3PP is a direct consequence of these constraints enabling responsible punishers to encourage self-interested opportunistic agents toward cooperation. Recall that with punishment reputation, opportunistic agents cooperate or defect depending on which action they expect to result in the better outcome, (i.e. payoff) for themselves. Thus the decision of an opportunistic agent to cooperate rather than defect occurs when

\[ \frac{c}{\rho} < P(R) - P(A), \]

where \( P(R) \) is the likelihood that the agent will be paired for punishment with a neighbor that will punish the agent responsibly, and \( P(A) \) is the likelihood that the agent will be paired with a neighbor that will punish the agent anti-socially. In a well-mixed population, \( P(R) \) and \( P(A) \) amount to the proportion of responsible punishers \( x_R \) and the proportion of anti-social punishers \( x_A \) in the population, respectively. Thus, with \( c=1 \) and \( \rho =3 \), an opportunistic agent would require \( x_R - x_A > 1/3 \). This means that, to induce cooperation in opportunistic agents, even with zero anti-social punishers, \( 1/3 \) of the entire population must be responsible punishers. In any sizable population, the likelihood that random exploration would lead to this ratio from a population of non-punishers is impossibly small. High strength-of-ties however alleviates this problem.

In a structured population, the quantities \( P(R) \) and \( P(A) \) in opportunistic agents' decision calculation depend on the punishment strategy of the neighbors. Specifically, for an agent \( v \) with neighborhood \( N(v) \) and degree \( d(v) \), if \( R^{N(v)} \) is the number of responsible punishers in \( N(v) \), then \( P(R) = \frac{R^{N(v)}}{d(v)} \). Similarly, \( P(A) = \frac{A^{N(v)}}{d(v)} \), thus \( P(R) - P(A) = \frac{(R^{N(v)} - A^{N(v)})}{d(v)} \), and we have the following version of Eqn 1. for agents on structured populations:

\[ \frac{c}{\rho} < \frac{(R^{N(v)} - A^{N(v)})}{d(v)} \]

The variable \( d(v) \) here, which represents the inverse of strength-of-ties, is crucial: a higher \( d(v) \) (lower strength-of-ties) means a lower probability of interacting with any given agent in the neighborhood. Since Eqn. 2 must hold for an opportunistic agent to cooperate, the higher \( d(v) \) the more responsible punishers must exist in order for self-interested, opportunistic agents to be induced toward cooperation. The lower \( d(v) \) however, the fewer responsible punishers are needed. Since punishing responsibly is only ultimately beneficial to the punishing individual if there are enough other similar punishers in the neighborhood, the lower \( d(v) \), the more favorable the conditions are for the evolution of responsible 3PP.

High mobility similarly hinders the evolution of responsible third-party punishment because, like low strength-of-ties, it renders the signaling of responsible punishers useless in promoting a sustained culture of cooperation in their neighborhood. Inducing cooperation in opportunistic agents requires the symbiotic existence of several responsible punishers in a neighborhood. When agents are highly mobile, it is difficult for punishers to maintain such localized coordination. Either needed fellow responsible punishers frequently move away, or non-cooperative agents frequently replace cooperative agents that have been induced as such in the neighborhood. Similar to conditions of low strength-of-ties, high mobility ultimately renders the cost of punishing responsibly fruitless, preventing the evolution of responsible 3PP.
Finally, we have conducted several experiments to further test the robustness of our findings. First, since population structure alone can aid cooperation under certain conditions, we also provide results for baseline experiments without 3PP in order to untangle the effects of 3PP from effects of population structure alone. Repeating our simulations with identical conditions but without the punishment phase shows that population structure alone does not account for the evolution of cooperation in the presented model. Even with high strength-of-ties and low mobility, cooperation does not emerge without 3PP. Hence the existence of 3PP is pivotal in the emergence of cooperation and increases overall payoff. Similarly, to unconfound the effects of 3PP and direct punishment, we have repeated these simulations with only direct punishers. Our results show that cooperation and responsible direct punishment cannot evolve alone in our model. This is because our model does not guarantee agents a chance to punish directly. When this is the case, 3PP is critically necessary for the evolution of responsible punishment and cooperation. Lastly, we have explored the evolution of 3PP when a separate trait for direct punishment can co-evolve. We find that while the existence of direct-only punishers decreases the overall prevalence of responsible 3PP, responsible 3PP remains necessary to induce cooperation. Hence responsible 3PP still evolves and promotes a high level of cooperation in the population as a whole under conditions of high social-structural constraint.

In summary, while the evolution of direct punishment has received considerable attention in the literature, the evolution of 3PP has not been well understood. Through a structured population model that implements variable degrees of social-structural constraint, we have found that high strength-of-ties and low mobility can provide a solution to the puzzle of the evolution of responsible 3PP. Responsible 3PP can evolve and induce cooperation with the help of punishment reputation when other mechanisms (e.g. population structure or direct punishment) alone fail to do so, but high strength-of-ties and low mobility are critical for this process.

Thwarting The Evolution of Conflict

In our final project for the grant, we sought to examine the conditions that can prevent the contagion of conflict. Nearly all major conflicts across the globe, both current and historical, are characterized by individuals defining themselves and others in terms of their group membership. Empirically, the existence of in-group favoring and out-group hostile behavior in humans is well established (Bernhard et al. 2006, Leroch & Hugh-Jones 2010). From an evolutionary perspective, numerous studies have shown how in populations comprised of various groups, group-biased behavior that discriminates or is hostile against out-groups evolves or emerges readily and dominantly (Hammond & Axelrod 2006, Choi & Bowles 2007, Antal et al. 2009, Garcia & van den Bergh 2011, Fu et al. 2012, McDonald et al. 2012, Hartshorn et al. 2013). Since humans are social beings who establish and define groups constantly, the development of out-group hostility and resulting group conflict might thus seem inevitable. Yet in a puzzling contrast, statistics have shown that violence and outgroup conflict have actually declined dramatically over the past few centuries of human civilization, suggesting out-group hostility is not inevitable after all (Pinker 2011a, 2011b). We asked: what factors might lead to such a decrease in conflict? Evolutionary game-theoretic models can shed light on this question by
exploring how various factors affect the emergence and maintenance of individuals’ behaviors relating to group conflict.

Our evolutionary game model builds on a prior model developed in Hammond and Axelrod’s pioneering work (2006) on the evolution of ethnocentrism, and used in Hartshorn, et al. (2013) In their model, agents had perceivable group tags, played one-shot Prisoner’s Dilemma games with their neighbors, and could behave differently toward in-group members than out-group members. Thus there were four possible strategies: Cooperate with both in-group and out-group members; Defect against both in-group and out-group members; Ethnocentric (cooperate with in-group members, defect against out-group members); Traitorous (defect against in-group members, cooperate with out-group members). Using their model with four different groups (or group tags), we have replicated their result showing that after a period in which Cooperative agents are briefly abundant, evolutionary pressure leads to a predominance of Ethnocentric agents. Defectors and Traitors never establish themselves (see the Supplementary Material for details).

Since the agents in that model conditioned their actions only on the group tags, they were in effect group-entitative. That leaves open the question whether there are conditions under which individual-entitative agents—agents that base their actions on knowledge of individuals per se rather than group tags—may be able to exist and perhaps even be favored by evolutionary pressures. Moreover, that model does not incorporate mobility. Research in cultural psychology has demonstrated large empirical differences in residential mobility around the globe with important psychological consequences (Long 1991, Angel 2000). Researchers have shown that in high-mobility contexts, individuals change relationships often; they form new relationships and sever unwanted relationships with great ease (Oishi et al. 2013, 2015). In such contexts, having a broad network of weak ties and being open toward strangers (with whom it might be valuable to form relationships) is highly adaptive. Indeed, Oishi, et al. (2015) observe that in highly mobile contexts, “since it is hard to keep track of behaviors of many strangers whom one meets, one needs to carefully avoid being associated with defectors or free-riders in order to exploit the greatest possible relational benefit” (p. 228). Thus, individuals are more likely to adopt strategies that try to evaluate the “trustworthiness and worth” (Oishi et al. 2015) of others in highly mobile contexts, i.e., adopt individual-entitative strategies. On the other hand, in low-mobility contexts, individuals have far fewer opportunities to form new relationships, and severing existing relationships can have extreme adverse effects such as being ostracized from one’s only social circle (Oishi et al. 2015), causing “the existential, social, and psychological death of the individual” (Landrine 1995, p. 755). Based on these theories we would predict that group-entitative behavior and associative ethnocentrism is adaptive in low mobility societies, yet it is maladaptive in high-mobility contexts, where individual-entitative strategies would be evolutionarily favored.

We ran extensive new evolutionary simulations, augmenting the prior model to include individual-entitative strategies and mobility; and our results show that the evolution of ethnocentrism is driven by low mobility. Indeed, our subsequent empirical analysis of archival data verifies that contexts with high residential mobility have less out-group hostility than those with low mobility.
In our evolutionary game model, agents are arranged on a lattice grid, and have Prisoner’s Dilemma interactions with their neighbors. Agents also have observable group tags and can distinguish between in-group and out-group members by observing these tags. Hence agents’ strategies can be conditioned on whether they are interacting with in-group or out-group members. Agents reproduce into empty neighboring locations with probability proportional to payoff achieved. Their offspring inherit their tags and strategies, with mutation (Traulsen et al. 2010, Allen et al. 2012). Locations become empty when agents die, and each agent has an equal chance to die at each time step. For additional details, see the Methods section.

Each agent can either be group-entitative or individual-entitative, and this is an inherited trait. We also allow agents to remember the actions of other agents they have encountered, and to have group-entitative and individual-entitative strategies based on memory. A group-entitative agent $i$ ignores individual identities. Its actions toward an agent $j$ depend only on its last encounter with anyone in $j$’s group. It has two possibly different strategies: one for in-groups and another for out-groups. Each of those strategies is one of the following: $AllC$ (always cooperate), $AllD$ (always defect), $TFT$ (Tit-for-Tat: play whatever action the opponent played in $i$’s last interaction with anyone from $j$’s group), or $OTFT$ (play the opposite of what $TFT$ would play). For details about $i$’s behavior during its first encounter with each group, see the Supplementary Material. An individual-entitative agent $i$ ignores other agents’ group tags; $i$’s action toward $j$ depends only on its last encounter specifically with $j$. Thus $i$ has one of the above four strategies, except that $TFT$ and $OTFT$ depend on $i$’s last interaction with $j$ specifically, rather than someone in $j$’s group. To model mobility, there is a probability $m$ with which, at the beginning of each iteration, an agent moves to a randomly chosen empty spot in the network. Thus a high value of $m$ represents a highly mobile population, while a low value of $m$ represents a population with low mobility. We vary $m$ from 0 to 0.08 in our experiments. It is important to note that a mobility probability of 0.08 is quite high: it means that on average, 8% of the population move to different locations on each iteration—a substantial amount of movement even for small values of $m$. At higher levels of mobility ($m>0.1$), cooperation breaks down in a society, and the majority of the population starts defecting—and thus is not representative of any stable society around the world (see the Supplementary Material for details).

Our results show strong support for the theory. Figure 1 shows our results after letting the populations evolve for 30,000 iterations. Without mobility (i.e., $m=0$), group-entitative agents comprise 75% of the population. These agents’ strategies are predominantly out-group hostile ($AllD$) and in-group cooperative ($AllC$). This is reasonably consistent with Hammond and Axelrod’s model (2006), but notice that even when $m=0$, individual-entitative agents comprise about 25% of the population. As mobility increases, the evolutionary pressures shift to favor individual-entitative agents. For $m>0.02$ they comprise about 80% of the population, and about 70% of them play $TFT$. Thus, the evolutionary dominance of group-entitative and ethnocentric strategies is thwarted by mobility.

The reason why low mobility favors group-entitative strategies while higher mobility favors individual-entitative strategies is related to the clustering of group members (Figure 2). With low mobility, groups tend to cluster together heavily; hence agents interact primarily with in-group members. Thus the ethnocentric strategy (i.e., group-entitativity with in-group cooperation and out-group-hostility) is effective and profitable in terms of payoff. Under higher
mobility, however, agents are less clustered by group membership, hence more likely to interact with out-group members, hence cannot rely on high payoffs from in-group interactions. Furthermore, group-entitative strategies are less effective because different group members are much less likely to have the same strategy. This favors the individual-entitative Tit-for-Tat (TFT) strategy.

To illustrate the evolutionary trajectories that led to the results reported in the main paper, Figures 3 and 4 show representative evolutionary trajectories for single simulation runs. In Figure 3, there is no mobility. Group-entitative agents quickly become a majority, and most of them are ethnocentric (in-group cooperative and out-group hostile). In Figure 4, the mobility probability is $m=0.05$. Individual-entitative agents evolve to become a majority, with most of those agents playing Tit-for-Tat (TFT).

**Empirical Analysis:** In order to complement these modeling efforts, we also gathered data to test the notion that mobility relates to lower ethnocentrism. We analyzed data from the U.S. Census Bureau (DDB Worldwide 1975-1998, Ren 2011) that provides measures of mobility in the U.S. 50 states (defined as the percentage of people born in the state of residence; reverse scored, with higher scores being reflective of higher mobility). We found that mobility was positively correlated with responses to the DDB Lifestyle survey (DDB Worldwide 1975-1998) “I am interested in the cultures of other countries” ($r = 0.614, p < .001$), and negatively correlated with DDB survey items regarding ethnocentrism (e.g., Americans should always buy American products, $r = -0.654, p < .001$; The government should restrict imported projects, $r = -0.578, p < .001$). In addition, states that have higher mobility also have higher openness, one of the big five personality dimensions, which is associated with breadth of experience and interest and interest in new ideas and other cultures ($r = 0.321, p = .023$) (John et al. 2008).
Figure 1: Proportions of actions and strategies as a function of mobility, after 30,000 iterations. Each data point is an average of 100 simulation runs. The plots show the proportions of (a) the group-entitative and individual-entitative agents, (b) the actions played by the agents, (c) the strategies of the individual-entitative agents, and (d) the in-group and (e) out-group strategies of the group-entitative agents.

Figure 2: Clustering coefficients after 30,000 iterations, for varying mobility values. Each data point is an average of 100 individual simulation runs. The degree of clustering decreases with mobility. For details on the metric used, see the Supplementary Material.
Figure 3: Single simulation run for 20000 generations with no mobility ($m=0$). (a) Proportions of group-entitative and individual-entitative agents. (b) Relative proportions of the individual-entitative agents’ strategies; Relative proportions of the group-entitative agents’ (c) ingroup and (d) outgroup strategies.
Figure 4: Single simulation run for 30000 generations with mobility probability $m=0.05$. (a) Proportions of group-entitative and individual-entitative agents. (b) Relative proportions of the individual-entitative agents’ strategies; Relative proportions of the group-entitative agents’ (c) ingroup and (d) outgroup strategies.

In all by integrating research on group conflict with human mobility (Nowak 2006, Oishi et al. 2007, Schug et al. 2009, Yamigishi and Suzuki 2009, Oishi 2010, Schug et al. 2010, West et al. 2011), we show for the first time that the evolution of ethnocentrism and group entitative behavior is thwarted by high mobility. As mobility is rapidly changing around the globe (Oishi et al. 2015), this work suggests that group conflict is not inevitable after all. This research is now under review at Nature Scientific Reports.

Publications, Awards, and Media


Gelfand, M.J. et al. (in press). Culture and getting to yes: Linguistic processes related to creativity in negotiation. *Journal of Organizational Behavior*

Fulmer, C. A, & Gelfand, M. J. (in press). Trust after violations: Are collectivists more or less forgiving?, *Journal of Trust Research*.

Nowak, Gelfand, Borkowski et al., (in press). The evolutionary basis on honor cultures. *Psychological Science*.

Workshop on Cultural Evolution, with top scholars from Biology, Anthropology, Computer Science, Linguistics, and Psychology, March 2016; *New Society for Cultural Evolution*.

Media: The Atlantic City Lab, Pacific Standard, Pys.org

Gelfand and/or Nau gave talks at Harvard, Yale, Georgetown, Duke, Kansas, University of Virginia, Washington University St. Louis, London Business School, and the National Academy of Sciences.
Other Honors and Awards to the PIs:

Gelfand, M. J. Won the new Anneliese Maier Research Award from the Alexander von Humboldt Foundation in Germany (Prize is $335,000) (http://www.newsdesk.umd.edu/uniini/release.cfm?ArticleID=2577), 2011.


1. Report Type
   Final Report

Primary Contact E-mail
Contact email if there is a problem with the report.
   mjgelfand@gmail.com

Primary Contact Phone Number
Contact phone number if there is a problem with the report
   2403726945

Organization / Institution name
   University of Maryland

Grant/Contract Title
The full title of the funded effort.
   CULTURE AND THE CONTAGION OF CONFLICT: SOCIAL SCIENCE AND COMPUTATIONAL APPROACHES

Grant/Contract Number
AFOSR assigned control number. It must begin with “FA9550” or “F49620” or “FA2386”.
   FA9550-12-1-0021

Principal Investigator Name
The full name of the principal investigator on the grant or contract.
   Michele J. Gelfand

Program Manager
The AFOSR Program Manager currently assigned to the award
   Dr. Benjamin Knott

Reporting Period Start Date
   11/15/2011

Reporting Period End Date
   05/14/2015

Abstract
We witness on a daily basis conflicts which spread from individuals quickly across groups, from the highly publicized incident that occurred when the Danish daily newspaper Jyllands-Postan published an article entitled “Muhammeds ansigt” which led to hundreds of protests and an escalation of violence, to the spread of conflict in Rwanda that caused the death of 800,000 individuals. This grant combined the use of social science and computational modeling techniques to illuminate the evolution of conflict contagion. We theorized that collectivism is a key driver of conflict contagion across social networks and across time through its impact on ingroup and outgroup entitativity. Our laboratory, field, and computational research showed strong support for the theory and illuminated important new scientific and practical insights. Our work was featured in top tier scientific outlets such as the Proceedings of the Royal Society B, Journal of Applied Psychology, Psychological Science, and Journal of Experimental Psychology, among others. A workshop on cultural evolution and conflict conducted through the grant has resulted in a new society for the study of cultural evolution which now has over 1000 members from many scientific disciplines.

Distribution Statement
This is block 12 on the SF298 form.
Archival Publications (published) during reporting period:


Moaddel M. & Gelfand, M. J. (forthcoming). Visions and perspectives in the study of human values in the
Middle East. Oxford University Press.


Fulmer, C. A, & Gelfand, M. J. (in press). Trust after violations: Are collectivists more or less forgiving?, Journal of Trust Research.


Media: The Atlantic City Lab, Pacific Standard, Pys.org

Gelfand and/or Nau gave talks at Harvard, Yale, Georgetown, Duke, University of Virginia, Washington University St. Louis, National Academy of Science

Other Honors and Awards to the PIs:

Gelfand, M. J. Won the new Anneliese Maier Research Award from the Alexander von Humboldt Foundation in Germany (Prize is $335,000) (http://www.newsdesk.umd.edu/uniini/release.cfm?ArticleID=2577), 2011.


Changes in research objectives (if any):
None

Change in AFOSR Program Manager, if any:
Dr Joseph Lyons and then Dr. Benjamin Knott.

Extensions granted or milestones slipped, if any:
One no-cost extension.

AFOSR LRIR Number
LRIR Title
Reporting Period
Laboratory Task Manager
Program Officer
Research Objectives
Technical Summary
Funding Summary by Cost Category (by FY, $K)
<table>
<thead>
<tr>
<th></th>
<th>Starting FY</th>
<th>FY+1</th>
<th>FY+2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salary</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equipment/Facilities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supplies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Report Document
Report Document - Text Analysis
Report Document - Text Analysis
Appendix Documents

2. Thank You

E-mail user

Aug 04, 2015 10:52:42 Success: Email Sent to: mjgelfand@gmail.com